The Sufficient Optimality Condition for Quantum Information Processing*

V P Belavkin
School of Mathematical Sciences, University of Nottingham,
Nottingham NG7 2RD
E-mail: vpb@maths.nott.ac.uk
A G Vantsyan

February 1, 2008

Abstract

The necessary and sufficient conditions of optimality of the decoding of quantum signals minimizing the Bayesian risk are generalized for the Shannon mutual information criteria. It is shown that for a linear channel with Gaussian boson noise these conditions are satisfied by coherent quasi-measurement of the canonical annihilation amplitudes in the received superposition.

1 Necessary and sufficient conditions of optimality

In [5, 3] dealing with optimization of the reception of quantum signals such as electromagnetic waves in the optical band, the search for the necessary conditions of optimality in the class of randomized strategies based on indirect measurements was our main concern. According to this universal approach, we shall specify the randomized strategies by the operator probability measures $\Pi(\mathrm{d}\beta)$ corresponding to quasi-measurements of certain noncommuting observables $b_j = \int \beta_j \Pi(\mathrm{d}\beta)$ such that

$$\Pi(\mathrm{d}\beta) \geq 0, \quad \int \Pi(\mathrm{d}\beta) = \hat{1}.$$

The equations derived in [5, 3] have the same form for both risk and information criteria of optimality:

$$(R(\beta) - \Lambda)\, \Pi(\mathrm{d}\beta) = 0, \quad \Lambda = \int R(\beta)\Pi(\mathrm{d}\beta), \tag{1.1}$$

*Originally published in: Radio Eng. Electron. Phys., **19** (7), p. 39, 1974. [trans. from Radiotekhnika i Electronika, 1974, **19**, 7, 1391–1395]

where $R(\beta) = \int c(\vartheta, \beta)\rho(\vartheta)P(\mathrm{d}\vartheta)$ is the "*a posteriori*" risk or nega-information operator in the Hilbert space of quantum states $\mathbb{H}$. Here $\rho(\vartheta)$ is the family of density operators describing the state of the quantum channel depending on the transmitted information ϑ with a prior distribution $P(\mathrm{d}\vartheta)$ and $c(\beta, \vartheta)$ is a given penalty function in Bayes case and the random information

$$i(\vartheta, \beta) = \ln \frac{P(\mathrm{d}\beta|\vartheta)}{\int P(\mathrm{d}\beta|\vartheta)P(\mathrm{d}\vartheta)}, \quad P(\mathrm{d}\beta|\vartheta) = \mathrm{Tr}\Pi(\mathrm{d}\beta)\rho(\vartheta) \tag{1.2}$$

with the opposite sign, $c(\vartheta, \beta) = -i(\vartheta, \beta)$, for the optimization criterion of maximum Shannon information

$$\mathsf{I}_{\beta,\vartheta} = \iint \ln \frac{P(\mathrm{d}\beta|\vartheta)}{P(\mathrm{d}\beta|\vartheta)P(\mathrm{d}\vartheta)} P(\mathrm{d}\beta|\vartheta)P(\mathrm{d}\vartheta).$$

It is obvious that operators $\Pi^{\mathrm{o}}(\mathrm{d}\beta)$ satisfying equation (1.1) are degenerate (provided $R(\beta) - \Lambda \neq 0$) and for each β have a range of values belonging to the zero eigensubspace of the difference $R(\beta) - \Lambda$. If operators $B(\beta) \equiv R(\beta) - \Lambda$ have a unique eigenvector φ_β for each β, corresponding to the zero eigenvalue, then the operator measure $\Pi^{\mathrm{o}}(\mathrm{d}\beta)$ is proportional to the projection operators $\Pi^{\mathrm{o}}(\mathrm{d}\beta) = \varphi^{\mathrm{o}}_\beta \varphi^{\mathrm{o}*}_\beta \mathrm{d}\beta$. In the general case when degeneration of the zero eigenvalue of operators $B(\beta)$ is possible: $B(\beta)\varphi_{\beta\nu} = 0$, $\nu \in N(\beta)$, each resolution of identity $\int \Pi(\mathrm{d}\beta) = \hat{1}$ satisfying Equation. (1.1) may be included in some more detailed resolution $\int \varphi_\gamma \varphi^*_\gamma \mathrm{d}\gamma = \hat{1}$ for $\gamma = (\beta, \nu)$. For different γ the vectors φ_γ describing the "elementary" measurements need not necessarily be orthogonal: $\varphi^*_\gamma \varphi_{\gamma^*} \neq 0$.

In Bayes case the sufficient conditions for optimality are very simple: the operators $\Pi^{\mathrm{o}}(\mathrm{d}\beta)$ satisfying Equation. (1.1) minimize the average risk

$$\mathsf{R} = \langle c(\vartheta, \beta) \rangle = \mathrm{Tr} \int R(\beta)\Pi(\mathrm{d}\beta) \tag{1.3}$$

if and only if the condition of nonnegative definiteness

$$B(\beta) \equiv R(\beta) - \int R(\beta')\Pi^{\mathrm{o}}(\mathrm{d}\beta') \geq 0 \tag{1.4}$$

is satisfied for all β. Actually, for any other operators measure $\Pi(\mathrm{d}\beta) \neq \Pi^{\mathrm{o}}(\mathrm{d}\beta)$ the difference

$$\mathsf{R} - \mathsf{R}^{\mathrm{o}} = \mathrm{Tr}\left[\int R(\beta)\Pi(\mathrm{d}\beta) - \int R(\beta)\Pi^{\mathrm{o}}(\mathrm{d}\beta)\right] = \mathrm{Tr}\int B(\beta)\Pi(\mathrm{d}\beta)$$

is nonnegative since it is the trace of a sum of products of nonnegative operators $B(\beta), \Pi(\mathrm{d}\beta)$.

Conditions (1.1) and (1.4) are applicable also for the optimization of the processing of quantum signals according to the maximum likelihood criterion. For this it is sufficient to consider that this criterion can be formally taken as Bayes criterion with uniform (unnormalized) *a priori* distribution $P(\mathrm{d}\vartheta) = \mathrm{d}\vartheta$

and a simple penalty function $c(\beta,\vartheta) = -\delta(\vartheta-\beta)$. This means that the a posteriori risk operator $R(\beta)$ should in this case be replaced by the density operator $\rho(\vartheta)$ at the estimate point $\vartheta = \beta$. We shall call the quantum strategies $\Pi^{\mathrm{o}}(\mathrm{d}\beta)$ satisfying conditions (1.1) and (1.4) for $R(\beta) = -\rho(\beta)$ optimum with respect to the maximum likelihood criterion. We shall give the solution of the problem of the discrimination of nonorthogonal signals for the following simplest case.

In [1] the concept of coherent processing a boson* signal $b = (b_\nu)$, i.e., of indirect linear measurement realized by measuring the superposition $b + a_0^*$, where a_0 is vacuum boson noise, was introduced. The question of the physical realization of this measurement was discussed at the Third All-Union Conference on the Physical Principles of Information Transmission by Laser Radiation. In particular it was shown [2] that coherent measurement of a narrowband optical signal can be realized by using an ideal heterodyne reception (ideal count of photons at different points of superposition of the received and the reference waves). The backward vacuum wave radiated by an ideally matched receiver into the communication line plays the role of noise a_0.

In [1] the quality of such processing was also defined from the maximum likelihood criterion for the case where b is the superposition of a coherent signal $\vartheta = (\vartheta_\nu)$ and a Gaussian boson noise a. The use of equations (1.1) and (1.4) and a suitable representation of the density operator makes it obvious that the processing described by the coherent projectors

$$\Pi(\mathrm{d}\beta) = |\beta\rangle\langle\beta|\mathrm{d}\mu(\beta), \quad \mathrm{d}\mu(\beta) = \prod_\nu \tfrac{1}{\pi}\mathrm{d}\,\mathrm{Re}\,\beta_\nu \mathrm{d}\,\mathrm{Im}\,\beta_\nu \tag{1.5}$$

is optimal. Such a suitable representation of the density operator $\rho(\vartheta)$ of the displaced Gaussian state $b = \vartheta + a$ is the representation in the form of the expression

$$\rho(\vartheta) = |L|^{-1} : \exp\left\{b-\vartheta)^\dagger L^{-1}(b-\vartheta)\right\} : \tag{1.6}$$

normally ordered with respect to the operators b^*, b. Here $L\big\|\langle\alpha_\nu\alpha_\nu^*\rangle\big\|$ is the correlation matrix of the noise a, the colon-brackets $: \cdot :$ denote normal order such that the operators a^* act to the left after the operators a, and $|L| \equiv \det L$.

Putting $R(\beta) = -\rho(\beta), \Lambda = -|L^{-1}|\hat{1}$, and considering the well-known [4] properties $\int |\beta\rangle\langle\beta|\mathrm{d}\mu(\beta) = 1$,

$$: p(b^*, b) : |\beta\rangle = p(b^*,\beta)\,|\beta\rangle$$

of the coherent vectors, we at once find that equation (1.1) has a unique solution coinciding with (1.5). The nonnegative definiteness of the operator

$$B(\beta) = |L^{-1}|\Big(1- : \exp\big\{-(b-\vartheta)^\dagger L^{-1}(b-\vartheta)\big\} : \Big)$$

*We recall that we are giving the name boson signal to the quantum signal described by the operators $\{\alpha_\nu, \alpha_\nu^*\}$ satisfying the commutation relations $\alpha_\nu\alpha_{\nu'} - \alpha_{\nu'}\alpha_\nu = 0$, $\alpha_\nu\alpha_{\nu'}^* - \alpha_{\nu'}^*\alpha_\nu = \delta_{\nu\nu'}$. In particular, the optical signal is described by the photon annihilation and creation operators a and a^* respectively.

is beyond doubt.

If the signals (ϑ_t) are known apart from the phase, the coherent processing is no longer optimum; however, it remains quasioptimal if the dimensionality of ϑ is large.

2 Local optimality according to information criterion

In the case of the information criterion it is difficult to get a global criterion of optimality of the solutions $\Pi^{\mathrm{o}}(\mathrm{d}\beta)$ of equation (1.1) minimizing the "Shannon" risk $\mathsf{R} = -\mathsf{I}_{\beta,\vartheta}$ in view of its nonlinear dependence on $\Pi(\mathrm{d}\beta)$ [the "penalty function" [3] also depends on $\Pi(\mathrm{d}\beta)$]. Therefore we give a differential criterion of optimality.

We shall restrict the discussion to operator probabilities of degenerate form $\Pi(\mathrm{d}\beta) = \varphi_\beta\varphi_\beta^*\mathrm{d}\beta$, where (φ_β) is the complete family of vectors $\int \varphi_\beta\varphi_\beta^*\mathrm{d}\beta = \hat{1}$. It can be shown that this is sufficient for the verification of local optimality $\delta^2\mathsf{R} > 0$ of the degenerate solutions $\Pi^{\mathrm{o}}(\mathrm{d}\beta) = \varphi_\beta^{\mathrm{o}}\varphi_\beta^{\mathrm{o}*}\mathrm{d}\beta$ of Equation. (1.1).

Making use of the dependence of the variations $\delta\varphi_\beta = \varphi_\beta - \varphi_\beta^{\mathrm{o}}$

$$\int (\varphi_\beta^{\mathrm{o}}\delta\varphi_\beta^* + \delta\varphi_\beta\varphi_\beta^{\mathrm{o}*} + \delta\varphi_\beta\delta\varphi_\beta^*)\,\mathrm{d}\beta = 0,$$

it is not difficult to find the increment $\Delta\mathsf{R} = \mathsf{R} - \mathsf{R}^{\mathrm{o}}$ of Shannon risk (1.3) (with the penalty function (1.2) depending on φ_β) at the stationary point $\varphi_\beta^{\mathrm{o}}$ with an accuracy up to second-order terms in $\delta\varphi_\beta$:

$$\begin{aligned}
&\Delta\mathsf{R} \simeq \iint \Big\{ c\,(\vartheta,\beta)\,\delta p(\beta|\vartheta) P(\mathrm{d}\vartheta) \\
&= \tfrac{1}{2}\Big[\big(\delta \ln p(\beta|\vartheta)\big)^2 - \big(\delta \ln p(\beta)\big)^2 \Big] p(\beta|\vartheta) p(\mathrm{d}\vartheta) \Big\} \\
&= \int \Big\{ \delta\varphi_\beta^* B(\beta)\delta\varphi_\beta = \tfrac{1}{2}\Big[\int (\delta\varphi_\beta^*\psi_\beta(\vartheta) + \psi_\beta^*(\vartheta)\delta\varphi_\beta)^2 p(\beta|\vartheta)P(\mathrm{d}\vartheta) \\
&- (\delta\varphi_\beta^*\psi_\beta + \psi_\beta^*\delta\varphi_\beta)^2 p(\beta)\Big]\Big\}\,\mathrm{d}\beta, \qquad (2.1)
\end{aligned}$$

where $B(\beta)$ is the difference (1.4),

$$\psi_\beta(\vartheta) = \frac{\rho(\vartheta)\varphi_\beta^{\mathrm{o}}}{p(\beta|\vartheta)}, \quad p(\beta|\vartheta) = \varphi_\beta^{\mathrm{o}*}\rho(\vartheta)\varphi_\beta^{\mathrm{o}},$$

and the vector $\psi_\beta = \rho\varphi_\beta^{\mathrm{o}}/p(\beta)$ $(p(\beta) = \int p(\beta|\vartheta)P(\mathrm{d}\vartheta))$ is the vector $\psi_\beta(\vartheta)$ averaged with the Bayessian posterior density

$$p(\vartheta|\beta) = p(\beta|\vartheta)p(\vartheta)/\int p(\beta|\vartheta)P(\mathrm{d}\vartheta).$$

A simple analysis of the positiveness $\delta^2\mathsf{R} > 0$ of the variation (2.1) of Shannon risk shows that in contrast to the Bayes case the nonnegativeness of the operators $B(\beta) \geq 0$ is necessary, but not sufficient for the local optimality of the solutions φ°_{β} of the equation $B(\beta)\varphi_{\beta} = 0$: the additional term in (2.1) (in square brackets) has the meaning of a posteriori variance of the real random quantity $2\,\mathrm{Re}\,\delta\varphi^*_{\beta}\psi_{\beta}(\vartheta)$ and is generally positive.

Let, for example, the density operator have the form (1.6) and the prior distribution $P(d\vartheta)$ be Gaussian in the multidimensional space of the information parameters $\vartheta = (\vartheta_\nu)$:

$$P(d\vartheta) = |S|^{-1}\exp\{-\vartheta^\dagger S^{-1}\vartheta\}\,d\mu(\vartheta), \quad d\mu(\vartheta) = \prod_\nu \tfrac{1}{\pi} d\,\mathrm{Re}\,\vartheta_\nu d\,\mathrm{Im}\,\vartheta_\nu.$$

We shall check the local optimality of the coherent solutions (1.5) of equation (1.1) in the boson Gaussian case according to the information criterion. As shown in [3], the coherent vectors (1.5) satisfy Equation. (1.1) and the operator $B(\beta)$ has a quadratic Gaussian form:

$$B(\beta) = (b-\beta)^\dagger H\rho(b-\beta),\ \rho = |L+S|^{-1}; e^{b^\dagger(L+S)^{-1}b} :,$$

where the matrix $H = L^{-1} - (S+L)^{-1}$ is not larger than unity in accordance with the inequalities $S \geq 0, L \geq 1, 0 \leq H \leq 1$. Considering the analytic dependence of the function

$$\psi_\beta(\vartheta) = \frac{\rho(\vartheta)|\beta\rangle}{\langle\beta|\rho(\vartheta)|\beta\rangle} = e^{-(b-\beta)^\dagger L^{-1}(\beta-\vartheta)}|\beta\rangle$$

on ϑ (i.e., the independence on ϑ^*) and carrying out conditional averaging over ϑ in (2.1),

$$\int \big(\delta\varphi^*_\beta\psi_\beta(\vartheta) + \psi^*_\beta(\vartheta)\delta\varphi_\beta\big)^2 p(\vartheta|\beta)d\mu(\vartheta)$$

with the density

$$p(\vartheta|\beta) = |M|\exp\big\{-(\vartheta - A\beta)^\dagger M(\vartheta - A\beta)\big\},$$

where $A = S(S+L^{-1}$, $M = S^{-1} + L^{-1}$, we find that in the Gaussian case the variation $\delta^2\mathsf{R}$ has the form

$$\delta^2\mathsf{R} = \int \delta\varphi^*_\beta\big(B(\beta) - D(\beta)\big)\delta\varphi_\beta\,d\mu(\beta),$$
$$d\mu(\beta) = \prod_\nu \tfrac{1}{\pi} d\,\mathrm{Re}\,\beta_\nu d\,\mathrm{Im}\,\beta_\nu,$$

where

$$D(\beta) =: p(b)\left[e^{-(b-\beta)^\dagger(1-H)(b-\beta)} - e^{-(b-\beta)^*(b-\beta)}\right] :\geq 0,$$
$$p(b) = |S+L|^{-1}\exp\big\{-b^\dagger(S+L)^{-1}b\big\}.$$

Thus, in order to prove the optimality of coherent quasi-measurement for the information criterion one should verify the operator inequality $B(\beta)-D(\beta)\geq 0$ or

$$p(\beta)e^{-b^{\dagger}(S+L)^{-1}\beta}:\Big[b^{\dagger}He^{-b^{\dagger}(S+L)^{-1}b}b - \Big(e^{-b^{\dagger}(H-1)b}-e^{-b^{\dagger}b}\Big)\Big]:e^{-\beta^{\dagger}(S+L)^{-1}b}\geq 0 \tag{2.2}$$

(here the change of variables $b-\beta\to b$ has been carried out). The operator occurring on the left-hand side of (2.2) has the structure $A^{*}_{\beta}:[\cdot]:A_{\beta}$ and is positive only if the operator in the square brackets is positive. Considering that the operator inequality

$$:e^{-b^{\dagger}(S+L)^{-1}b}:\geq:e^{-b^{\dagger}(1-H)b}:$$

is satisfied by virtue of the matrix inequality $(S+L)^{-1}=L^{-1}H\leq 1-H$, we find that inequality (2.2) is satisfied if

$$:b^{\dagger}He^{b^{\dagger}(H-1)b}b:\geq:e^{b^{\dagger}(H-1)b}-e^{-b^{\dagger}b}:. \tag{2.3}$$

This inequality (2.3) becomes obvious in the diagonal representation in the occupation numbers n_{ν}:

$$\prod_{\nu}h_{\nu}^{n_{\nu}}\sum_{\nu}n_{\nu}\geq\prod_{\nu}h_{\nu}^{n_{\nu}}\quad\text{for}\quad\sum_{\nu}n_{\nu}\neq 0,$$

where h_{ν} are the eigenvalues of the matrix $H=L^{-1}-(S+L)^{-1}$. For $\sum_{\nu}n_{\nu}=0$ both the left- and right-hand sides of inequality (2.3) vanish.

References

[1] V.P. Belavkin, *Radiotekhnika i Elektronika*, 1972, **17**, No. 12, 2533 [Radio Eng. Electron. Phys., **17** No. 12 (1972)].

[2] V.P. Belavkin, Coherent measurement of optical signals, Proc. of the Third All-Union Conference on the Physical Principles of Recording and Processing Inofrmation by Laser Radiation, Kiev, 1973, I, p. 7.

[3] V.P. Belavkin and R.L. Stratonovich, *Radiotekhnika i Elektronika,* 1973, **19**, 9, 1839 [Radio Eng. Electron. Phys., **19** 9 (1973)].

[4] J. Klauder and E. Sudershan, Fundamentals of Quantum Optics (in Russian transl.), Mir Press, 1970.

[5] R.L. Stratonovich, The quantum generalization of optimal statistical estimation and hypothesis testing, *J. of Stochastics*, 1973, **1**, 87–126.